Weekly Report

2018.0625-2018.0701

1.This Week

Power Grid Deep Learning Paper

1.We find a problem of our generated data: we construct the declaration file of the simulation program to output all nodes' info of the grid but the program only returns 30% nodes' info. We're contacting Xu and Huang to solve this problem (we have been communicating since the monday of this week, but Xu is on a business trip and the problem is still not solved with the information he gives us. He says when he returns next week, he will help us have a look at the program.).

2.Since the data problem is hanging there, we can only progressed other branches of this paper:

- With the data we now have, we test simple learning algorithms such as SVM, decision tree. The accuracy is only about 10%, indicating that neural networks can actually catch more information.
- We apply simple random sampling and uncertainty-based sampling of each instance to simulate incomplete training data. (When we increases the accuracy of the model trained on the complete training data, we plan to randomly sample 25% nodes in each instance and repeat this process to generate incomplete training data sets for each instance and use these data sets to train the model to try to get a good performance.)

SQC Paper

1. Zongzhuang is realizing the CUSUM algorithm with the real case power grid data (data of 36 nodes in the power grid). He is still not finished and we set the deadline of this part to 7.06

Others

1.Read the book machine learning written by Zhou Zhihua and learning basic theories (especially mathematical backgrounds) about machine learning and deep learning. (This week about active learning.)

2.revise the waveline paper. Except for the paper revision, we need to re-do the evaluation part to get quantitative evaluation results.

WaveLines Revision Plan:

changes to be made:

【done】1.make clear the definitions of power gris simulation terms like transient simulation stc..(make a form)

2.clearly define all patterns mentioned in this paper and discuss how existing works distinguish these patterns in the related work.

3.revise the evaluation part: add quantitative evaluation of how accurate wavelines can help to find patterns(effenctiveness), how long will take(efficiency) and complete understanding (comprehensiveness).
【done】4.explain the data preprocessing into a whole continuous process.
5.related work: explain why methods used to visualize pairwise variables mentioned are not used in this paper.(use this to guide the alternative designs discussion.)
6.use less words but add a figure to explain the section waveline design trade-off.
7.explain more carefully why we use a bulb metaphor.

problem:
1.reviewer 3 suggest that expert feedbacks are now informative after the first revision but it is still difficult to judge.
2.reviewer 3 challenges the generability of the approach: we explain that it can be applied to multi-variate time series in a pairwise way, but the reviewer regard it as can be only applied to pairwise variables.

Paper Reading

1.RCLens: Interactive Rare Category Exploration and Identification

This paper proposes a visual analytics system to explore and identify user-guided rare categories. It is a typical VAST paper and includes minor visualization novelty. There are two things I learn from this paper: 1) the most common way to create novel visualization designs is to combine two or more classic visualizations together; 2) The analysis pipeline in this paper is clear and rich, making all views closely tighted but not to add views to rich the paper.
The algorithm to detect rare categories in this paper is active learning based algorithm. The query strategy they used is defined by themselves but based on uncertainty sampling.

2.Improving Generalization with Active Learning

This paper introduces the basic ideas of active learning. It presents a theory of selective sampling and considered active learning as selective sampling (not random sampling.) This theory has been tested with a neural network (SG).

3.Active Learning with Statistical Models

This paper proposes statistical optimal ways as query strategies to select training data. It proves that old optimal selection techniques has achieved an excellent performance with feedforward neural networks and their statistical model can achieve a better performance for the statistically-based learning algorithms like mixtures of guassians and locally weighted regression. The greatest advantage of statistical models is that it can be more efficient and statistically accurate.

2.Progress

Work	Deadline	Progress
Power grid paper with Deeping learning	-	1.Improving the model accuracy. 2.Deal with the data problem.
SQC Paper	-	About to started.
WaveLine revision	ASAP	Follow the revision plan to revise the paper.